Solución de los ejercicios de autocomprobación

1. Debemos contrastar que los datos proceden de una distribución normal. Como tenemos 12 datos (menos de 30) no podemos utilizar el contraste de la χ^2 y realizamos el contraste de Kolmogorov-Smirnov.

   Como no tenemos los parámetros, debemos estimarlos por el método de máxima verosimilitud. Sabemos que en una normal, el estimador de la media es la media muestral y el de la varianza la varianza muestral (coinciden los EMV con los estimadores por el método de los momentos). Así, si X es la variable aleatoria “Contenido en grasas de la leche que recibe CLP”,

$$\hat{\mu} = \bar{X} = 4{,}073; \quad \hat{\sigma} = \sqrt{m_2} = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} = 0{,}357$$

   Por tanto, la hipótesis nula es $X \sim N(4{,}073, 0{,}357)$.

   Para calcular la función de distribución teórica, sabemos que

$$X \sim N(\mu, \sigma) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

   Se obtiene la siguiente tabla.

x_h	$F_n(x_h)$	$z_h = \frac{x_h - 4{,}073}{0{,}357}$	$F(x_h)$	$\lvert F_n(x_h) - F(x_h)\rvert$	$\lvert F_n(x_{h-1}) - F(x_h)\rvert$
3.45	0.083	-1.745	0.0405	0.0425	0.0405
3.54	0.167	-1.49	0.0681	0.0989	0.0149
3.82	0.25	-0.71	0.2389	0.0111	0.0719
3.89	0.333	-0.51	0.305	0.028	0.055
3.9	0.417	-0.48	0.3156	0.1014	0.0174
3.97	0.5	-0.29	0.3859	0.1141	0.0311
4.19	0.583	0.33	0.6293	0.0463	0.1293
4.25	0.667	0.5	0.6915	0.0245	0.1085
4.36	0.75	0.8	0.7881	0.0381	0.1211
4.43	0.833	1	0.8413	0.0083	0.0913
4.53	0.917	1.28	0.8997	0.0173	0.0667
4.55	1	1.34	0.9099	0.0901	0.0071

   Como $\{\lvert F_n(x_h) - F(x_h)\rvert\} = 0{,}1141$ y $\{\lvert F_n(x_{h-1}) - F(x_h)\rvert\} = 0{,}1293$, tenemos que $D_{12} = 0{,}1293$. Buscando en la tabla de Kolmogorov-Smirnov para la normal con media y varianza desconocidas, $D_{12,0{,}05} = 0{,}242$. Como $D_{12} = 0{,}1293 < D_{12,0{,}05} = 0{,}242$, no existe evidencia muestral para rechazar H_0 con un nivel de significación del 5 %, es decir, no podemos rechazar la distribución normal con $\mu = 4{,}073$ y $\sigma = 0{,}357$.

2. Sea X = *importe pagado en cada factura (en euros)*. Debemos comprobar que esta variable aleatoria sigue una distribución $N(\mu, \sigma)$, utilizando el contraste de la χ^2 de Pearson. Los estimadores de máxima verosimilitud de μ y σ son la media muestral y la desviación típica muestral, respectivamente. Así, la hipótesis nula a contrastar es

$$H_0 : X \sim N(90, 8)$$

Construimos la tabla:

Clase i	O_i	p_i	$E_i = np_i$	$\frac{(O_i - E_i)^2}{E_i}$
≤ 75	4	0.0304	6.08	0.7116
$(75,\ 81]$	26	0.0999	19.98	1.8138
$(81, 87]$	42	0.2236	44.72	0.1654
$(87, 93]$	69	0.2922	58.44	1.9082
$(93, 99]$	36	0.2236	44.72	1.7003
$(99, 105]$	12	0.0999	19.98	3.1872
≥ 105	11	0.0304	6.08	3.9813
Total	n= 200	1		13.4679

Para el cálculo de las probabilidades se ha utilizado la tabla de la función de distribución de la $N(0, 1)$. Si Z es una v.a. $N(0, 1)$ y F es su función de distribución, se obtienen de la siguiente forma:

$p_1 = P(X \leq 75) = P\left(Z \leq \dfrac{75 - 90}{8}\right) = P(Z \leq -1{,}875) = F(-1{,}875) = P(Z \geq 1{,}875) = 1 - F(1{,}875) = 1 - 0{,}9696 = 0{,}0304$

$p_2 = P(X \in (75, 81]) = P(Z \in (-1{,}875, -1{,}125]) = F(-1{,}125) - F(-1{,}875) = 1 - F(1{,}125) - 0{,}0304 = 1 - 0{,}8697 - 0{,}0304 = 0{,}0999$

...

Si H_0 es cierta,

$$D^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-r-1=7-2-1=4}$$

El valor observado para la discrepancia es $\hat{D}^2 = 13{,}4679$. Como no nos especifican un nivel de significación, debemos calcular el nivel crítico o p-valor:

$$p = P(D^2 \geq 13{,}4679 / H_0 \text{ cierta}) = P(\chi^2_4 \geq 13{,}4679)$$

Por interpolación lineal obtenemos $p \simeq 0{,}00937$, que es inferior a 0.01 y por tanto concluimos que existe evidencia para rechazar la hipótesis nula, es decir, rechazamos que la muestra procede de una distribución $N(90, 8)$.

3. Dada la variable aleatoria X = *tiempo (en minutos) hasta que termina la clase*, debemos contrastar si sigue una distribución uniforme en el intervalo [2,3]. Así, la hipótesis nula es $H_0 : X \sim U(2, 3)$.

Realizamos el contraste de Kolmogorov-Smirnov. El contraste de la χ^2 de Pearson no se puede utilizar porque tenemos menos de 30 observaciones.

La función de distribución de una uniforme en el intervalo $[a, b]$ es

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx = \int_{a}^{x} \frac{1}{b-a}dx = \frac{x-a}{b-a}$$

si $a \leq x \leq b$.

Con lo que la función de distribución teórica es $F(x) = x - 2$.

Se obtiene la siguiente tabla. Al construirla hay tener en cuenta que hay dos datos que se repiten.

x_h	$F_n(x_h)$	$F(x_h)$	$\lvert F_n(x_h) - F(x_h)\rvert$	$\lvert F_n(x_{h-1}) - F(x_h)\rvert$	$D_{10}(x_h)$
2.11	$\frac{1}{10} = 0.1$	0.11	0.01	0.11	0.11
2.13	0.2	0.13	0.07	0.03	0.07
2.24	0.3	0.24	0.06	0.04	0.06
2.49	0.5	0.49	0.01	0.19	0.19
2.58	0.7	0.58	0.12	0.08	0.12
2.67	0.8	0.67	0.13	0.03	0.13
2.7	0.9	0.7	0.2	0.1	0.2
2.82	1	0.82	0.18	0.08	0.18

donde $D_{10}(x_h) = \{|F_n(x_h) - F(x_h)|, |F_n(x_{h-1}) - F(x_h)|\}$, con lo que

$$D_{10} = D_{10}(x_h) = 0{,}2$$

Buscando en la tabla de Kolmogorov-Smirnov, $D_{10,0{,}05} = 0{,}4093$. Como $D_{10} = 0{,}2 < D_{10,0{,}05} = 0{,}4093$, no existe evidencia muestral para rechazar la distribución uniforme, con un nivel de significación del $5\,\%$.

4. Sea X la variable aleatoria *número de pilas defectuosas entre las cuatro pilas de una linterna*. Debemos contrastar la hipótesis nula:

$$H_0 : X \sim Binomial(4, p)$$

La probabilidad p de pila defectuosa se debe estimar utilizando los datos muestrales. El estimador de máxima verosimilitud de p es

$$\hat{p} = \frac{\sum X_i}{4n} = \frac{\bar{X}}{4} = \frac{61 + 34 \times 2 + 13 \times 3 + 3 \times 4}{150 \times 4} = 0{,}3$$

Así,

$$P(X = x) = \binom{4}{x} 0{,}3^x 0{,}7^{4-x}, \quad x = 0,\, 1,\, 2,\, 3,\, 4$$

Debemos utilizar el contraste de la χ^2 de Pearson. Construimos la tabla:

x_i	O_i	p_i	$E_i = 150p_i$	$(O_i - E_i)^2/E_i$
0	39	0.2401	36.015	0.2474
1	61	0.4116	61.74	0.008869
2	34	0.2646	39.69	0.815724
3	13	0.0756	11.34	0.242998
4	3	0.0081	1.215	2.6224

Observemos que la última clase tiene un valor para la frecuencia esperada inferior a 5, con lo que debe juntarse con la clase anterior. Sin embargo, si hacemos tal agrupación nos quedaríamos con sólo 4 clases. Como no es posible cumplir todas las reglas el resultado del test no va a ser muy fiable.

$$\hat{D}^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 3{,}93739$$

Si H_0 es cierta, $D^2 \sim \chi^2_{k-r-1=5-1-1=3}$.

Como $\hat{D}^2 = 3{,}93739 < \chi^2_{3,0{,}05} = 7{,}815$, no existe evidencia suficiente para rechazar H_0.

Observación: Si agrupamos las dos últimas clases (valores 3 y 4), $\hat{D}^2 = 2{,}017273$ que es menor que $\chi^2_{2,0{,}05} = 5{,}991$ y llegamos a la misma conclusión.